What Metarepresentation is For

What metarepresentation is for

It is widely accepted that many creatures do have minds, but most people agree as well that not many creatures, if any but humans, know that they have minds. Humans not only have beliefs, desires, etc, but they also think about these states all the time. This ability to direct the mind away from the environment and onto itself is not just any old ability either. Many people believe that it is crucially important in explaining why human minds seem to be so much more powerful than the ones of any other creature we know of.[1] It has for example been argued that it allows humans to take their minds offline, to break the direct link between perception and action (see e.g. Prinz (2003) dual representation). Knowing that the lion you contemplate is no more than a thought and not the real thing allows you to think in peace about what would be the best thing to do if you really were to meet one. If you are not able to distinguish between thoughts and the real thing, then the only thing that thinking about a lion will prompt is a fast escape up the next tree. In addition, recognising thoughts as thoughts allows you to understand that they can be false, and this knowledge should allow you to make much better predictions about your fellow beings. As well, in the case of yourself, it will allow you to not take at face value what you believe now, but to re-examine the steps that led you to a certain belief, and to generally regiment the way you do your thinking.

Thus, it is clear that thinking about thinking is a crucially important ability of humans. What is a lot less clear is what it actually consists in, and whether there really is only one ability that could be referred to as thinking about thinking. Indeed, if one looks more closely at the literature, one realises very quickly that this is very doubtful. In this chapter, I set out to contrast two ways of what it could mean that humans know that they have minds. In particular, I want to focus on the question of whether we need the ability to metarepresent in order to think about our minds. It will turn out that, at least on some senses of thinking about thinking, this is not a requirement. This might seem surprising, as metarepresentation is supposed to be nothing else than representing representations as such. Indeed, one of the most (if not the most) basic characteristics of the mind is that they (what does "they" refer to?) represent[2]. So, if you have no concept of what a representation is, how can you think about things which are at heart representational things?

[1] Really, all the way back via Strawson to Kant.
[2] Or in less cog sci language, have aboutness or intentionality.

The trick that makes it nevertheless possible to think about your thoughts without understanding the nature of representation is that the representations in human minds are not only mental states, but also have contents. Many people have argued that it is possible to think about contents as contents without understanding the nature of mental representational states.[3] Here I want to look at the attempt to use language as the tool to make this odd sounding claim plausible[4]. Specifically, I will start by looking at the idea of the 'self regulating mind' as defended by Pettit/McGeer (2002) and explore with them the idea that one way of explaining the difference between human and other minds is that, in contrast to other minds, human minds are self regulating and that this self regulation is achieved by thinking about content as content in lingual form. However, Pettit/McGeer do not clearly distinguish between different forms of self regulation. It is here where this chapter departs from their account and explores which forms of self regulation do require the ability to metarepresent.

The self regulating mind

Pettit and McGeer claim in their article "The self regulating mind" that what sets human minds apart from the minds of other creatures is that they are self regulating and not merely routinized. They claim that the crucial difference between the two is that self regulating minds can do things intentionally to control their beliefs which routinized minds can't. Self regulating minds do this by using the unique properties of language. Humans, unlike any other creature, represent the world not only in their beliefs, but also in the sentences that they use to express those beliefs.[5]

It is the properties of this second representational system that, according to Pettit/McGeer make human minds special. On the one hand, language is special because it is very closely related to beliefs. If someone sincerely asserts p then she thereby expresses her belief that p.

[3] One way to achieve this would obviously be introspection. Many people were very fond of this route, but obviously, there are as well many famous problems associated with it. Ultimately, I think that introspection is a dead end, but it is not relevant to defend this view here in depth. See for example Eric Schwitzgebels (2008) view for a contemporary critique of introspection.

[4] People who have tried to use language for that claim are e.g. Dennett (1998) or Clark (2006).

[5] This is not supposed to be a simple on/off thing. In fact, Clark (2006) discusses cases from the animal literature where the presence of material symbols enables monkeys to significantly enhance their abilities. Nevertheless, it is obviously equally true that no other animal is anywhere near the same ball park as humans when it comes to mastery of language.

But on the other hand, language does have some special properties as well which set it apart from beliefs. The first property that Pettit/McGeer discuss is the fact that language allows for what they call 'content attention'. Routinized minds obviously have contents as well, but as Pettit/McGeer write, they are blind to them. They, as it were, look through them into the world. Language, on the other hand, solidifies the content into what Andy Clark calls 'material symbols' (2006).[6] These material symbols provide a new set of representational objects for the mind to work with.[7] Pettit/McGeer write that this enables attention to contents as such. But it seems more adequate to say that it is not only content that language delivers. Language delivers as well a completely new set of vehicles that bear the contents. Obviously, they only do so in a derived way, but they are content-bearing vehicles nevertheless. The huge advantage of these vehicles, in contrast to the vehicles of beliefs, is that they are visible in the world (in the case of the spoken word this will obviously be replaced by audible) and manipulable for the cognizer. This allows the cognizer to direct her attention to all the different properties of the content stored in these visible/audible vehicles. As Pettit/McGeer point out, one of the most important consequences of this is that the cognizer now can learn much more easily about the properties that constrain propositions and thereby learn to build up a network of propositions that respect the identified constraints.

The difference that is most important for Pettit/McGeer, however, is the fact that in contrast to believing, asserting is something that we can do voluntarily. Whether or not someone believes that p is normally not up to them (sentence unfinished?). It depends on their (what does "their" refer to?) evaluation of the evidence for p, but whether or not somebody asserts that p is entirely under their voluntary control. Obviously, what exactly voluntary control is, is contentious, but here it is enough to say that assertion, but not belief, is voluntary in a very similar sense to lifting one's arm. If the agent decides to do it, then it will normally happen. Mc Geer/Pettit believe that this difference in intentional control is at the heart of the difference between routinized minds and self regulating minds. It allows humans to voluntarily implement constraints that improve their thinking. They can now voluntarily give themselves greater exposure to relevant evidence for their deliberations, they can intentionally

[6] Another important name here is clearly Dennett (e.g. 1998), who developed a similar account to Clark, but sees language as even more transformative.

[7] So really what language does is give the cognizer a second representational system. Once we understand this as what language does, it becomes clear that there could be other ways to achieve this second and / or offline representational stream. Even though I find the idea that language provides this very plausible, I do want to remain neutral about alternative possibilities. What really matters for me is to clarify the relationship between this second representational stream (however constructed), the mastery of folk psychology (or mental state concepts), and the intentional control of the mental.

control their reasoning process, by e.g. repeating steps, reminding themselves of important considerations, and so on. In addition, intentional control even allows humans to overcome very strong natural evaluative tendencies. McGeer/Pettit discuss the example of the pilot who can learn to overcome the tendency to form a belief because of proprioceptive signals and to intentionally use the instrument panel to guide her behaviour, even if it recommends actions which are in sharp contrast to proprioceptive demands.[8]

These are clearly all very important abilities and it seems fair to say that they very plausibly could justify the distinction between non human routinized minds and human self regulating minds.[9] However, there is one move in the paper which demands closer attention: a very surprising twist about halfway through the McGeer/Pettit paper. Up until then, they describe how language enables attention to content and the identification of constraints for this content. But then they prepare the next step in their argument (which is about how humans can intentionally implement content) by saying that humans ascribe beliefs and desires to themselves and that they understand themselves as intentional systems (288). This clearly seems to be about ascribing mental states with specific contents to oneself, rather than ascribing pure content to oneself (it seems incomprehensible, at least to me, to ascribe pure belief contents, but not states). It is one thing to attend to material symbols without attending to psychological states – in fact, this is, as discussed, the beauty of language, i.e. that it does give a new set of content-bearing vehicles, about which we can think without having to have a clue about how these things might work in the human mind. But it is quite another thing to ascribe beliefs to oneself, without understanding that one thereby ascribes a psychological state. As beliefs are presumably psychological states, this seems simply contradictory.

So perhaps understanding mental states is actually far more important for self regulating minds than Pettit/McGeer at first seem to make out. That Pettit/McGeer might think so is borne out as well by the way their paper progresses. They discuss the importance of folk psychology in many contexts, and again it seems obvious that folk psychology is about ascribing mental states, rather than about attending to and manipulating pure content.
But more importantly, it is not only the fact that Pettit/McGeer seem to think that the ascription of mental states plays a crucial role for the self regulating mind, the basic structure

[8] This fits very nicely with the mental muscle view of the will as defended by Baumeister (2008) and Holton (2009).
[9] This is obviously not to say that there might not be many other equally plausible ways of accounting for the differences.

of intentional control of minds seems to require an understanding of mental states as states (please clarify/simplify this sentence).
In order to intentionally manipulate anything, it seems necessary that you have some kind of hold on what it is that you are manipulating and in the case of manipulating the mind, that is clearly a mental state.[10] If it were the case, however, that all intentional control of the mental as described by Pettit/McGeer requires metarepresentation, then we would now have found what we really need metarepresentation for, and it would be very important for a very large chunk of human thinking indeed.

But do we really need to be able to ascribe folk psychological states to ourselves in order to intentionally manipulate our mental lives? A first reason to be doubtful about this comes from animal research. In Michael Beran's lab (2009), for example, they could show that monkeys can use intentional strategies in order to master delayed gratification tasks. When offered the choice between a small reward now or a bigger reward later, the monkeys started playing with a toy in order to distract themselves from the urge to take the smaller reward now.[11] This looks very much like the monkeys used intentional behaviour in order to prevent a re-evaluation of their belief that the benefits of a large reward later are bigger than the ones of a small reward now, which would have happened if they had focussed on the small reward. It actually looks remarkably similar to the pilot case described by Pettit/McGeer. Crucially, however, nobody seems to suspect that the monkeys have mastered metarepresentation or possess a folk psychology and were ascribing mental states to themselves. So clearly, the monkeys are acting intentionally to bring about a desired result and get the bigger reward, but they are clearly not aware of the mechanism that makes it the case that the strategy that they employ is successful.

But where one might still have some doubts in the monkey case as to whether these animals do not have something like a folk psychology, there are much more banal cases where intentional action prevents or allows the re-evaluation of a given situation. Take e.g. the turning of its head by a nervous mouse in order to see whether the buzzard it was watching fearfully earlier on is still on the tree at a safe distance where it had been before. Turning its head clearly allows the mouse to update, or if you prefer, to regulate its beliefs, but it would

[10] Pamela Hieronymi (2009) e.g. calls all intentional mental action managerial control and describes this form of control as intentionally manipulating mental objects. She also refers to managerial control as attitude directed control, which really says it all.
[11] The toy was under normal circumstances considered boring by the monkeys, so they really only played with it when it was used for distraction purposes.

seem ludicrous to suggest that it does so because it is aware that its belief about the location of the predator is by now possibly false.

This is the solution to our puzzle then: It is quite possible to manipulate mental states by acting intentionally, even if you don't know what mental states are, if the manipulation of mental states is not what you intend to achieve but is connected to (or is e.g. the cause of) you successfully reaching a first order goal.

Taking this insight on board, we can now return to the human case and ask ourselves whether we really need metarepresentation in order to go to the library or in order to rehearse an argument? Do we really need to understand ourselves as psychological creatures in order to do these things? The answer is that no, we do not: It seems quite enough if we understand that propositions can be true or false, and that we can find evidence for their truth or falsity by going to the library or rehearsing an argument. Obviously understanding the truth of a proposition is quite different from being able to employ a distracting strategy, as in the case of the monkeys, or from the even more basic head-turning behaviour of the mouse. It is in one sense not first order, but thinking about thinking, because, thanks to language, humans but not monkeys[12] or mice understand something about the features of content (that it e.g. can be right or wrong, as well supported by the evidence, etc.). In a different sense of thinking about thinking, however, it is just as much first order as in the case of the monkeys or mice. Going to the library to find better evidence for the proposition you are interested in does not require you to understand that this proposition is the content of a psychological state. You do not have to be able to ascribe such states to yourself and you do not have to be a folk psychology user. Instead, you can simply concentrate on the properties of the content, which are visible to you thanks to their embeddedness in the material symbols of the language vehicle.[13]

So we actually do not need to be able to employ the intentional stance[14] in order to intentionally regulate our minds, but obviously, as long as we have not mastered mental state

[12] This is probably not an on/off phenomenon and it is therefore very likely that monkeys do already possess a few material symbols. See e.g. Clark (2006).

[13] In this respect, Hieronymi seems to have got it wrong, if she thinks that managerial (i.e. intentional) control of mental states must be attitude directed – at least if attitude directed means that the attitude is what the agent is aware of targeting in the action.

[14] We have to be careful here, because it is not clear whether the intentional stance as developed by Dennett (1987) really is a psychological stance. If we give it a behaviourist reading, then it might actually be reducible to the kind of thing that language enables. On such a reading, however, the intentional stance is then unsuitable to understand humans as psychological creatures.

concepts we will not be aware of the fact that that is what we are doing. At the same time, it could still be true that it is thinking about thinking that makes self regulation in the human case categorically different from the intentional control that mice and monkeys have over their minds. Language gives humans a huge range of new objects that they can intentionally manipulate (think about) which are not available to animals. These objects that human intentions can be about will be the propositions we have in the form of lingual material symbols. So even though I disagree about the importance of folk psychology for being able to use language as a self regulation tool, it might obviously still be the case that the intentional control which is enabled by language is what makes all the difference between us and animals. On the other hand, it might not be. It might be as well the other features of language discussed here that make all the difference.[15] Whatever the right answer, what matters in this context is that there are intentional forms of self regulation which do not require that the agent understand anything about mental states. This leaves us now with the question of whether this understanding is simply irrelevant in the context of the intentional regulation of the mind. It is this question to which we will now turn.

What metarepresentation is for

Mindreading

In the last section I argued that folk psychology is not necessary for intentionally regulating one's mind. This ability may, combined with language, be crucial for explaining the difference between animal and human minds, but attributing mental states to oneself is not necessarily part of this form of self regulation by thinking about thinking (don't you mean, rather: attributing mental states to oneself by thinking about thinking is not necessarily part of this form of self regulation)

From this conclusion we now move naturally into the second part of the paper: if self regulation in the sense described in the last section does not require metarepresentation, what do we need it for then? Which are the situations where it does matter whether we know that the material symbols we are manipulating derive their intentionality from the beliefs of people? The obvious candidate here is the ascription of false beliefs. As long as we do not

[15] I take it that Pettit/Mc Geer would probably be on one side here, Mele (2009) and Holton (2009) somewhere in the middle, and Moran – even though I do think there is some ambiguity in the notion of deliberation as employed by Moran (2001) – and Hieronymi (where it sometimes feels as if deliberation actually does include non aware intentional control), and Strawson (2003) on the other.

understand that people's beliefs can deviate from the rational norm, it seems very difficult to imagine how one can predict that someone will act or even that someone has a false belief. Obviously, this was exactly the reasoning that led to the debate defining the false belief task (Wimmer&Perner, 1983).

Nowadays, however, there are some significant worries about the need for a folk psychological theory in solving false belief tasks. The problem is that false belief task-like competencies have been found not only in preschoolers, but also in toddlers (e.g. Buttelmann (2009) and infants (e.g. Onishi (2005)). If we do not want to claim that these (in the case of the infants: prelingual!) children possess mental state concepts, then it must be possible to solve false belief tasks without such concepts. But if it is possible to do that, then there is now a heavy burden of proof on defenders of the theory theory, who claim that children use theory in the later explicit tasks (see e.g. Perner (2011) for Povinelli and Voh's challenge and Perner's reply to it).[16]

But whether or not this challenge can be successfully answered, there is an additional reason to be sceptical about the importance of possessing mental state concepts for our discussion here. This is because our topic is not mind reading, but the function of metarepresentation in metacognition. In other words, our question is: what is the function of theory-based mind reading for self? But if it is the case that mindreading is important in the case of other minds, because others might not share our beliefs and we need the psychological theory to help us to understand this psychological fact so that we can make better predictions, then it seems clear that the case of our own minds will be quite different. After all, as Moore (refce please, maybe footnote about Moore's paradox useful for non philosophers) has famously pointed out, it does sound very odd indeed, if one asserts a proposition, but denies that one believes it.

What is the use of mindreading for self?

[16] I do not discuss the also folk psychology heavy narrative alternative to theory theory here (please describe briefly the narrative theory). It seems to me though that if narrative theory is used for making claims about how it is that we acquire the ability to solve false belief tasks, then the same worry as for theory theory applies. However, often the narrative alternative in many variations seems to be about doubting that this is what folk psychology is for. For a discussion of the e.g mindshaping hypothesis, which is related to the narrative theory, see the section 'Self as other'.

One of the most discussed worries for theory theory always was that it does not respect the intuitively obvious difference between self knowledge and knowledge of other minds.[17] It is intuitively quite plausible that I might need a piece of theory in order to figure out what you are thinking, but it does seem very counterintuitive that the same should be true when I want to know what I am thinking. Obviously, the literature within philosophy that tries to explain what it is that makes knowledge of our own minds special is vast, but one explanation has been derived by Pettit/McGeer directly from the idea of the self regulating mind (please revise this sentence, the "but" does not seem to contrast with anything) . According to the so-called agency theory of self knowledge, what is special about self knowledge is that in order to find out what you believe you do not have to do any psychology, but simply deliberate about the first order question. Once you have found your answer to the first order question you have as well answered your belief question. One has special authority about one's own beliefs, because by answering the question one creates (or at least makes visible) the relevant state.[18] Importantly, in this context, if self knowledge is created because it automatically flows from the deliberations of the agent, then it cannot normally go wrong. It is quite different from normal observational knowledge and therefore it seems difficult to see why we would need any psychological theory to improve it.

Self as other

This rule (please remind the reader what "this rule" is) does have exceptions, however – when we are interested in our psychology (do you mind in our own mind?) not because of the contents of its states, but because of their nature as states. This could be for two quite different reasons. On the one hand, we might think that certain mental states have a positive (or negative) value as states rather than as content, (example useful) while on the other hand, it might be the case that we are interested in mental states as states, because this allows us to see ourselves as psychological beings rather than as rational agents (example). This allows us to take into account the fact that our psychology might change and we might one day think something is rational which we now would consider plainly irrational, or that there might be

[17] For a very good introduction to the debate between theory theory and simulation theory see Davies (1995).
[18] Pettit/McGeer build here on the work of Richard Moran. However, it has to be mentioned that there is an important ambiguity in her (his?) work. Deliberation Moran style is not an intentional affair, but in line with what Hieronymi calls evaluative control, whereas in Pettit/McGeer it is at least partly intentional (or managerial) in Hieronymi's terminology.

psychological traits in us that influence our thinking and acting even though they do not appear in our rational justifications.

A very nice example of the former is Pascal's wager. Pascal felt that he had a very good reason to believe in God, even if his existence was very unlikely. He argued that the belief would not do much harm, even if it were wrong, but that the consequences would be dire if one did not believe and it turned out that God existed. One problem for Pascal was that it is not clear how knowing that it would be good to have the belief in God would help in any way in acquiring it. Pascal's solution to this problem ingeniously employed the theoretical stance. He argued that it is quite possible to acquire the belief, even against one's rational evaluative tendencies, by simply conditioning oneself in the right way. So Pascal recommended going to mass and praying regularly and over time the sheer force of habit would produce the belief that seemed out of reach by rational means. Pascal uses psychological knowledge in order to achieve the odd state of asserting a proposition and at the same time asserting confidently that he will not believe it very soon (don't you mean: he will believe it very soon) (because he knows he will have conditioned himself to do so).

A similar case is Kavka's famous toxin puzzle. In it, an agent gets offered a large sum of money, if at midnight she has the intention to drink a mildly unpleasant toxin the next day. The puzzle arises, because it seems very difficult for a rational agent to collect the reward. This is because as the agent gets nothing for actually drinking the toxin, but only for having the intention, as she has absolutely no reason to drink it when the time comes. As she knows this, it becomes impossible for her to *intend* to drink it, because in order to intend something the agent has to be settled on doing it, and this agent knows that there is no reason for actually drinking it. As puzzling as this story is, it obviously only works as long as the agent is not allowed to form the intention by non-rational means.

Future directed self control

In his book *Reasons and Persons*, Derek Parfit (1984) includes the example of a rich young communist who ardently believes in material equality, but who also knows that statistically most rich young communists turn into rich old conservatives. This communist faces the dilemma of whether he should give away the money now, and therefore be true to the ideals he now holds, or keep it, thereby maximizing his chances of doing what he will

come to think is the right thing to do and which, given the assumption that he will be a rich conservative for longer than he is a rich communist, is what maximizes doing what he thinks is the right thing over his lifetime. The example is wonderful, because it illustrates how incredibly deep the chasm between our first order thinking and our psychological perspective on ourselves can be. The rich conservative that he will be is worrying for the communist, not because he thinks that by the time he is old he will be too weak-willed for his ideals, but because he believes the evidence that suggests that older people are more likely to genuinely judge that it is morally wrong to give up the property that has been in the family for generations. He knows that, even though he is connected to his future self by means of a continuous and slowly changing psychology, it is quite likely that the values that make his decisions in an important sense his as an agent on many accounts of free agency will have been replaced by ones he now deplores.[19] Of all the deep issues on ethics and personal identity that the example raises, what matters for us here is that it illustrates incredibly well what our psychological perspective on ourselves adds to first order reasoning. It allows us to see our future selves as somebody else[20] and to realise that even our deepest-felt convictions are possibly merely transient psychological states.

Most importantly, however, realising that our first order evaluations may change enables us to have a completely new level of self-control. As long as the agent does not have psychological knowledge, it will be nigh on impossible for her to conceive that something which she very strongly believes to be true now could be judged by her to be false by tomorrow. As she can sincerely see no evidence that would render the proposition in question false, it becomes very difficult, if not impossible, for her to comprehend that she might nevertheless perceive it to be false tomorrow.(Last two sentences very similar)

This, then, is the difference between the intentional control of the mental that is enabled by having a new set of material symbols which one can intentionally manipulate and the control that is dependent on the understanding of the fact that one is a psychological creature. Only the second allows for targeted interventions that aim at the states rather than the contents. (the states, however, are here considered because of their respective contents).

[19] In Parfit's architecture, this example serves to show that we should not overestimate the importance of doing the right thing across a life time.
[20] We have to be careful here, because it is by no means certain that we really most of the time use the psychological stance when interpreting others.

The targeted interventions that one can only use competently and flexibly if one understands oneself as a psychological being are the many old self-control strategies that humans have so successfully invented over the course of their history. They range from providing external constraints on undesirable action options (like tying yourself to the mast to prevent yourself from jumping to your death or giving your car keys to the landlord in order to prevent yourself from achieving the same result by driving) to directly influencing your psychology. (maybe at this point it would be useful to contrast these interventions with Beran's monkeys using toys to distract themselves from attractive stimulus). The latter can be achieved by providing relevant input for the machinery (remind yourself what a hangover feels like) or even by direct tampering with the machinery (a couple of drinks will make you less shy).

Importantly, as we have seen with the monkeys, there are only very few things in self control which you cannot do at all if you do not understand mental states (Pascal-like cases are the only ones that spring to mind), but understanding that you are a psychological creature opens up a whole new league of flexibility in employing these tools. (the two parts of the sentence seem to offer contradictory evaluations: few things or a whole new league? Maybe add: "seems" or "prima facie")

Mindshaping

Before we come to the end, it seems very important to put the argument in this chapter in relation to a related but different recent development in the mindreading literature. Based on loosely Dennettian roots, thinkers like Dan Hutto (2008) or Tad Zawidsky (2008) have recently made a strong case for the idea that folk psychology might not be for mindreading, but for mindshaping. The core of the idea is that human development is scaffolded by the constant ascription of propositional attitudes to children, even at an age when these children do not really understand these terms. These attributions act then like self fulfilling prophecies. They provide a normative standard that the children try to meet.

This seems to me to be a powerful account of the use of folk psychology, but if it is true, one might think that it undermines the position defended here. For one, it seems to show that folk psychology can be used to shape minds, even at a time when children do not yet understand metarepresentation, and secondly it seems to show that theory based mindreading is not necessary for shaping minds later either, because the mindshaping happens simply

because of the practice of ascribing mental states to each other, which then act as normative markers. All of this seems to be possible without the need for theory.

In reply to this it should be pointed out that this chapter is mainly interested in the *intentional* self regulation of the human mind. Mindshaping does not really have an awful lot to say about that (argument needed: it seems to me that mindshaping needs to finally resort to intentional self-regulation to achieve any change at all in an individual's mind). Secondly, the argument about the use of language developed in this chapter is supposed to show that language-related thinking about thinking gives massive cognitive advantages which are quite unrelated to folk psychology.[21] In this respect, then, language-based thinking about thinking is much wider than the scaffolding provided by ascribing folk psychology terms to children. Finally, it is again the issue of flexibility that singles out theory-based self regulation as discussed in the second half of the chapter. Once the agent is aware of what it is that she is doing, self regulation can become much more efficient and flexible than in a case where mindshaping happens only in the unreflected normative marker way as decribed by Hutto or Zawidsky. All in all, then, mindshaping seems highly compatible with the ideas defended here.

Conclusion

In this chapter we examined two different ways in which thinking about thinking enables intentional self regulation. On the one hand, we looked at the way in which language provides us with a new set of representational objects for the mind to manipulate. The great thing about these objects, in contrast to the vehicles of beliefs, is that they can be intentionally manipulated in a very straightforward way. Agents can choose what to assert and this means that it becomes much easier to attend to (rehearse, refine, etc.) contents. However, it became clear that this ability does not require one to attribute or even understand the nature of mental states. It is not the mastery of folk psychology which enables this control. In the second half of the chapter we then asked ourselves whether there is some form of intentional self control which does require the mastery of folk psychology. We discussed two forms of control where this seems to be the case. Obviously, an understanding of mental states is required, if what we are interested in is the acquisition of such states as states. However, these cases do seem rather rare, so it was the second form on which we concentrated. We argued that an

[21] And they are a lot wider even than the behaviourist intentional stance reading that Zawidsky espouses (could you please clarify for the reader?).

understanding of mental states is crucial for forms of intentional self-control where we need to take into account the fact that our psychology can change over time. In these cases, as long as we do not have an understanding of ourselves as psychological beings, we will struggle to effectively achieve our self-control aims. Folk psychology might not be what separates humans from animals, but it seems likely that if, as most people seem to assume, self control is crucial for autonomy, an understanding of folk psychology is crucial for an autonomous agent.

Bibliography:

Michael Beran et al (2009). Delay of gratification by chimpanzees (Pan troglodytes) in working and waiting situations. Behavioural Processes, 80, 117-121.

Baumeister, R. (2008). Free will, consciousness and cultural animals. Are we free? J. Baer, OUP**:** 65-85.

Clark, A. (2006) Material Symbols. *Philosophical Psychology, 19 (3), 1-17.*

Davies, M., & Stone, T. (Eds.). (1995). *Folk psychology*. Blackwell.

Dennett, D. (1987). *The intentional stance*. MIT Press.

Dennett, D (1998). Reflections on language and mind. In Carruthers, P., & Boucher, J. (Eds.), *Language and thought: interdisciplinary themes*. Cambridge University Press.

Hieroymi, P. (2009). Two kinds of mental agency. In O'Brien, L. & Soteriou, M. (Eds.), *Mental Actions.* Oxford, OUP.

Holton, R. (2009). *Willing, wanting, waiting*. Oxford ; New York, Oxford University Press.

Hutto, D (2008) *Folk Psychological Narratives. The sociocultural basis of understanding reasons.* Cambridge MA, MIT Press

Mele, A. (2009) Mental actions a case study. In O'Brien, L. & Soteriou, M. (Eds.), *Mental Actions.* Oxford, OUP.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258.

Perner, J. (to appear in 2011). Theory of Mind—an unintelligent design: From behaviour to teleology and perspective. In Leslie, A.M., German, T. (Eds.), *Handbook of 'Theory of Mind'*. Lawrence Erlbaum Associates.

Pettit, P., & McGeer, V. (2002). The self regulating mind. *Language and Communication*, 281- 299.

Prinz, W. (2003). Emerging selves. Representational foundations of subjectivity. *Consciousness and Cognition,12, 515-528.*

Strawson, G. (2003). Mental Ballistics Or The Involuntariness of Spontaneity. Meeting of the Aristotelian Society, London.

Schwitzgebel, E. (2008), The unreliability of naive introspection. *Philosophical Review, 117* 245-273.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103-128.

Zawidzki, T. (2008). The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations* 11(3): 193 – 210